

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Masure, H. Robert
Rosenow, Carsten I.
Tuomanen, Elaine
Wizemann, Theresa M.

(ii) TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
ANTI-PNEUMOCOCCAL VACCINES

(iii) NUMBER OF SEQUENCES: 25

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: David A. Jackson, Esq.
(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
(C) CITY: Hackensack
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07601

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Jackson Esq., David A.
(B) REGISTRATION NUMBER: 26,742
(C) REFERENCE/DOCKET NUMBER: 600-1-158 ..

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 201-487-5800

(B) TELEFAX: 201-343-1684

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP112

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Glu Asn Glu Gly Ser Thr Gln Ala Ala Thr Ser Ser Asn Met Ala
1 5 10 15

Lys Thr Glu His Arg Lys Ala Ala Lys Gln Val Val Asp Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP90

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Arg Glu Phe Ser Leu Glu Lys Thr Arg
1 5 10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa Arg Glu Phe Ser Leu Glu Lys Thr Arg Asn Ile Gly Ile Met Ala
1 5 10 15

His Val Asp Ala Gly Lys Thr

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

(2) INFORMATION FOR SEQ ID NO:5:

(ii) MOLECULE TYPE: peptide
(A) DESCRIPTION: cBP78

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gln Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Ala Val Ala
1 5 10 15
Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro Glu
20 25 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP70

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Xaa	Xaa	Xaa	Glu	Val	Ala	Lys	Xaa	Ser	Gln	Asp	Thr	Thr	Thr	Ala	Ser
1				5				10						15	

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP60

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa	Asn	Glu	Arg	Val	Lys	Ile	Val	Ala	Thr	Leu	Gly	Pro	Ala	Val	Glu
1				5				10						15	

Gly Arg Gly

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP50 pep

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Xaa Ile Ile Xaa Xaa Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly
1 5 10 15

Asn Pro

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP112-Int1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

SECRET

[illegible]

SECRET

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Trp Gln Pro Pro Arg Ala Arg Ile
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: 50 KDa protein

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Thr Asp Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *B. subtilis*

(xi) SEQUENCE DESCRIPTION: SEQ ID,NO:13:

Ile Val Asp Val Arg Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA.

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCGCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT

(2) INFORMATION FOR SEQ ID NO:15:

(A) LENGTH: 210 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

(2) INFORMATION FOR SEQ ID NO:16:

(A) LENGTH: 70 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID.NO:16:

Ser Arg Gly Asn Pro Thr Leu Glu Val Glu Val Tyr Thr¹ Glu Ser Gly
1 5 10 15

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Glu Ala
50 55 60

Ile Ile Gly Tyr Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iv) ANTI-SENSE: NO

(A) ORGANISM: B. subtilis

Ser Arg Gly Asn Pro Thr Val Glu Val Glu Val Tyr Thr Glu Thr Gly
1 5 10 15

Ala Phe Gly Arg Ala Leu Val Pro Ser Gly Ala Ser Thr Gly Gln Tyr
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Asp Arg Tyr Leu Gly Lys
35 40 45

Gly Val Leu Thr Ala Val Asn Asn Val Asn Glu Ile Ile Ala Pro Glu
50 55 60

Leu Leu Gly Phe Asp Val
65 70

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1511 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCCGCGGCA	ACCCAACACT	TGAAGTAGAA	GTTTACACTG	AATCAGGTGC	TTTCGGACGT	60
GGTATGGTTC	CATCAGGAGC	TTCTACTGGT	GAACACGAAG	CAGTTGAACT	TCGCGACGGT	120
GACAAATCTC	GTTACGGTGG	TCTTGGTACA	CAAAAAGCTG	TTGACAACGT	AAACAACATC	180
ATTGCTGATT	CTATCATTTG	CTACGATGTA	CGTGATCAAC	AAGCTATTGA	CCGTGCTATG	240
ATCGCACTTG	ACGGTACTCC	TAACAAAGGT	AAATTGGGTG	CGAATGCAAT	CCTCGGTGTG	300
TCTATCGCTG	TAGCTCGTGC	TGCTGCTGAC	TACCTTGAAA	TCCCACCTTA	CAGCTACCTT	360
GGTGGATTCA	ACACTAAAGT	TCTTCCAACT	CCAATGATGA	ACATCATCAA	CGGTGGTTCT	420
CACTCTGACG	CTCCAATCGC	TTTCCAAGAG	TTCATGATCT	TGCCAGTTGG	TGCGCCAACA	480
TTTAAAGAAA	CCCTTCGTTA	CGGTGCTGAA	ATCTTCCACG	CTCTTAAGAA	AATCCTTAAA	540
TCACGTGGTT	TGGAAACTGC	CGTAGGTGAC	GAAGGTGGAT	TCGCTCCTCG	TTTCGAAGGA	600
ACTGAAGATG	GTGTTGAAAC	TATCCTTGCT	GCGATTGAAG	CTGCTGGATA	TGTACCAGGT	660
AAAGACGTAT	TTCTCGGATT	TGACTGTGCT	TCATCAGAAT	TCTACGATAA	AGAACGTAAA	720
GTTTACGACT	ACACTAAATT	TGAAGGCGAA	GGTGCTGCTG	TTCGTACATC	TGCAGAACAA	780
ATCGACTACC	TTGAAGAATT	GGTTAACAAA	TACCCAATCA	TCACTATTGA	AGATGGTATG	840
GATGAAAACG	ACTGGGATGG	TTGGAAAGCT	CTTACTGAAC	GTCTTGGTAA	GAAAGTACAA	900
CTTGTTGGTG	ACGACTTCTT	CGTAACAAAC	ACTGACTACC	TTGCACGTGG	TATCCAAAAA	960
GGTGCTGCTA	ACTCAATCCT	TATCAAAGTT	AACCAAATCG	GTACTCTTAC	TGAAACTTTT	1020
GAAGCTATCG	AAATGGCTAA	AGAAGCTGGT	TACACTGCTG	TTGTATCACA	CCGTTCAGGT	1080

(2) INFORMATION FOR SEQ ID NO:19:

(A) LENGTH: 420 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(iv) ANTI-SENSE: NO

(A) ORGANISM: *S. pneumoniae*

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His

30

Leu Gly Phe Asp Cys Ala Ser Ser Glu Phe Tyr Asp Lys Glu Arg Lys
225 230 235 240

Asp Leu Lys Lys
420

(A) LENGTH: 429 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAAAAAGGTA GCAGAAGCTG AGAAGAAGGT TGAAGAAGCT GAGAAAAAAG CCAAGGATCA	60
AAAAGAAGAA GATCGCCGTA ACTACCCAAC CAATACTTAC AAAACGCTTG ACCTTGAAAT	120
TGCTGAGTCC GATGTGAAAG TTAAAGAAGC GGAGCTTGAA CTAGTAAAAG AGGAAGCTAA	180
GGAACCTCGA GACGAGGAAA AAATTAAGCA AGCAAAAGCG AAAGTTGAGA GTAAAAAAGC	240
TGAGGCTACA AGGTTAGAAA ACATCAAGAC AGATCGTAAA AAAGCAGAAG AAGAAGCTAA	300
ACGAAAAGCA GCAGAAGAAG ATAAAGTTAA AGAAAAACCA GCTGAACAAC CACAACCAGC	360
GCCGGTTACT CAACCAGAAA AACCAGCTCC AAAACCAGAG AAGCCAGCTG AACCAACCAA	420
AGCAGAAAA	429

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

10040 22262800

Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro Lys Ala Glu
130 135 140

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: other nucleic acid

- (iii) HYPOTHETICAL: NO

GCTCTTNCTC GATGTCTCNG TNGCCAT

27

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: other nucleic acid

- (iii) HYPOTHETICAL: NO

AGC⁺ATAGTCT TCTTCGACTT GTTGATCATC

30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAAGCAGCC ACTTCTTCTA ATATGGCAAA GACAGAACAT AGGAAAGCTG CTAAACAAGC 60
GTCGATGAAT ATATAGAAAA AATGATGAGG GAGATTCAAC TAGATAGAAG AAAACATACC 120
CAAAATGTCG CCTTAAACAT AAAGTTGAGC GCAATTAAAA CGAAGTATTT GCGTGAATTA 180
AATGTTTTAG AAGAGAAGTC GAAAGATGAG TTGCCGTCAG AAATAAAAGC AAAGTTAGAC 240
GCAGCTTTTG AGAAGTTTAA AAAAGATACA TTGAAACCAG GAGAAAAGGT AGCAGAAGCT 300
AAGAAGAAGG TTGAAGAAGC TAAGAAAAAA GCCGAGGATC AAAAAGAAGA AGATCGTCGT 360
AACTACCCAA CCAATACTTA CAAAACGCTT GAACTTGAAA TTGCTGAGTT CGATGTGAAA 420
GTTAAAGAAG CGGAGCTTGA ACTAGTAAAA GAGGAAGCTA AAGAATCTCG AAACGAGGGC 480
ACAATTAAGC AAGCAAAAAGA GAAAGTTGAG AGTAAAAAAG CTGAGGCTAC AAGGTTAGAA 540
AACATCAAGA CAGATCGTAA AAAAGCAGAA GAAGAAGCTA AACGAAAAGC AGATGCTAAG 600
TTGAAGGAAG CTAATGTAGC GACTTCAGAT CAAGGTAAAC CAAAGGGGCG GGCAAAACGA 660
GGAGTTCCTG GAGAGCTAGC AACACCTGAT AAAAAAGAAA ATGATGCGAA GTCTTCAGAT 720

TCTAGCGTAG	GTGAAGAAAC	TCTTCCAAGC	TCATCCCTGA	AATCAGGAAA	AAAGGTAGCA	780
GAAGCTGAGA	AGAAGGTTGA	AGAAGCTGAG	AAAAAAGCCA	AGGATCAAAA	AGAAGAAGAT	840
CGCCGTAACT	ACCCAACCAA	TACTTACAAA	ACGCTTGACC	TTGAAATTGC	TGAGTCCGAT	900
GTGAAAGTTA	AAGAAGCGGA	GCTTGAAC TA	GTAAAAGAGG	AAGCTAAGGA	ACCTCGAGAC	960
GAGGAAAAAA	TTAAGCAAGC	AAAAGCGAAA	GTTGAGAGTA	AAAAAGCTGA	GGCTACAAGG	1020
TTAGAAAACA	TCAAGACAGA	TCGTAAAAAA	GCAGAAGAAG	AAGCTAAACG	AAAAGCAGCA	1080
GAAGAAGATA	AAGTTAAAGA	AAAACCAGCT	GAACAACCAC	AACCAGCGCC	GGCTACTCAA	1140
CCAGAAAAAC	CAGCTCCAAA	ACCAGAGAAG	CCAGCTGAAC	AACCAAAAGC	AGAAAAAACA	1200
GATGATCAAC	AAGCTGAAGA	AGACTATGCT	CGTAGATCAG	AAGAAGAATA	TAATCGCTTG	1260
ACTCAACAGC	AACCGCCAAA	AACTGAAAAA	CCAGCACAAAC	CATCTACTCC	AAAAACAGGC	1320
TGGAAACAAG	AAAACGGTAT	GTGGTACTTC	TACAATACTG	ATGGTTCAAT	GGCAACAGGA	1380
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1440
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTTCAAT	GGCAACAGGA	1500
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAC	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1560
TGGCTCCAAT	ACAATGGTTC	ATGGTACTAC	CTAAACAGCA	ATGGCGCTAT	GGCGACAGGA	1620
TGGCTCCAAT	ACAATGGCTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1680
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1740
TGGCTCCAAT	ACAACGGTTC	ATGGTATTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGT	1800
TGGGTGAAAG	ATGGAGATAC	CTGGTACTAT	CTTGAAGCAT	CAGGTGCTAT	GAAAGCAAGC	1860
CAATGGTTCA	AAGTATCAGA	TAAATGGTAC	TATGTCAATG	GCTCAGGTGC	CCTTGCAGTC	1920
AACACAAC TG	TAGATGGCTA	TGGAGTCAAT	GCCAATGGTG	AATGGGTAAA	CTAAACCTAA	1980

TATAACTAGT TAATACTGAC TTCCTGTAAG AACTTTTTTAA AGTATTCCTT ACAAATACCA 2040
TATCCTTTCA GTAGATAATA TACCCTTGTA GGAAGTTTAG ATTAAAAAAT AACTCTGTAA 2100
TCTCTAGCCG GATTTATAGC GCTAGAGACT ACGGAGTTTT TTTGATGAGG AAAGAATGGC 2160
GGCATTCAAG AGACTCTTTA AGAGAGTTAC GGGTTTTTAA CTATTAAGCC TTCTCCAATT 2220
GCAAGAGGCT TCAATCTCTG CTAGGGTGCT AGCTTGCGAA ATGGCTCCAC GGAGTTTGGC 2280
AGCGCCAGAT GTTCCACGGA GATAGTGAGG AGCGAGGCCG CGGAATTCAC GAACTGCGAC 2340
GTTTTCTCCT TTGAGGTAA TCAATCGTTT CAAGTGTTCG TAGGCGATCT TCATCTTGTC 2400
TTCAAAGGTC AAATCAGGTA GGATTTCTCC TGTTTCAAAG TTTATGGTGG CCCTGGTTGA 2460
AG 2462

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met Met Arg Glu Ile Gln Leu Asp Arg Arg Lys His Thr Gln Asn Val
1 5 10 15

Ala Leu Asn Ile Lys Leu Ser Ala Ile Lys Thr Lys Tyr Leu Arg Glu

30

Ser Leu Lys Ser Gly Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu
225 230 235 240

Ala Met Ala Thr Gly Trp Leu Gln Asn Asn Gly Ser Trp Tyr Tyr Leu
450 455 460

Asn Gly Glu Trp Val Asn Glx
625 630

(1) GENERAL INFORMATION:

- (i) APPLICANT: Masure, H. Robert
Rosenow, Carsten I.
Tuomanen, Elaine
Wizemann, Theresa M.
- (ii) TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
ANTI-PNEUMOCOCCAL VACCINES
- (iii) NUMBER OF SEQUENCES: 25
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: David A. Jackson, Esq.
(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
(C) CITY: Hackensack
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07601
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Jackson Esq., David A.

(C) REFERENCE/DOCKET NUMBER: 600-1-158 ..

(A) TELEPHONE: 201-487-5800

(B) TELEFAX: 201-343-1684

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(A) DESCRIPTION: CBP112

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Glu Asn Glu Gly Ser Thr Gln Ala Ala Thr Ser Ser Asn Met Ala
1 5 10 15

Lys Thr Glu His Arg Lys Ala Ala Lys Gln Val Val Asp Glu
20 25 30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP90

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala	Arg	Glu	Phe	Ser	Leu	Glu	Lys	Thr	Arg
1				5					10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP84

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Xaa	Arg	Glu	Phe	Ser	Leu	Glu	Lys	Thr	Arg	Asn	Ile	Gly	Ile	Met	Ala
1					5				10					15	

His Val Asp Ala Gly Lys Thr

20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP80

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Xaa	Lys	Xaa	Xaa	Trp	Gln	Xaa	Lys	Gln	Tyr	Leu	Lys	Glu	Asp	Gly	Ser
1				5				10						15	

Gln	Ala	Ala	Asn	Glu	Xaa	Val	Phe	Asp	Thr	Ala
			20					25		

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
 (A) DESCRIPTION: cBP78

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gln	Lys	Ile	Ile	Gly	Ile	Asp	Leu	Gly	Thr	Thr	Asn	Ser	Ala	Val	Ala
1				5					10					15	
Val	Leu	Glu	Gly	Thr	Glu	Ser	Lys	Ile	Ile	Ala	Asn	Pro	Glu		
			20					25					30		

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
 (A) DESCRIPTION: cBP70

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Xaa	Xaa	Xaa	Glu	Val	Ala	Lys	Xaa	Ser	Gln	Asp	Thr	Thr	Thr	Ala	Ser
1					5					10				15	

200002220000

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP60

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Xaa Asn Glu Arg Val Lys Ile Val Ala Thr Leu Gly Pro Ala Val Glu
1 5 10 15

Gly Arg Gly

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: cBP50 pep

(iii) HYPOTHETICAL: NO.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

(2) INFORMATION FOR SEQ ID NO:9:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP112-Int1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr Tyr Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: cBP112-Int2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Xaa	Asp	Asp	Gln	Gln	Ala	Glu	Glu	Asp	Tyr	Ala
1				5						10

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Trp	Gln	Pro	Pro	Arg	Ala	Arg	Ile
1				5			

T02070-2822300

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: peptide

- (v) FRAGMENT TYPE: N-terminal

(A) ORGANISM: S. pneumoniae

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ile Thr Asp Val Tyr Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: 50 KDa protein

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ile Val Asp Val Arg Ala Arg Glu Val Leu Asp Ser Arg Gly Asn Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 210 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCGCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT

GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAACT TCGCGACGGT 120

GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAGCTG TTGACAACGT AAACAACATC 180

ATTGCTGAGG CCATCATTGG CTACGATGTA 210

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCCCGCGGCA ACCCAACAGT TGAAGTTGAA GTATATACAG AAACAGGAGC TTTCGGCCGC 60

GCATTAGTGC CAAGCGGAGC TTCTACAGGT CAATACGAAG CGGTTGAGCT TCGTGACGGC 120

GACAAAGACC GTTACCTTGG AAAAGGCGTG TTAACAGCTG TTAACAACGT AAACGAAATC 180

ATTGCTCCAG AGCTTCTTGG CTTTGATGTA 210

(2) INFORMATION FOR SEQ ID NO:16:

- (A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser Arg Gly Asn Pro Thr Leu Glu Val Glu Val Tyr Thr Glu Ser Gly
1 5 10 15

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly. Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Glu Ala
50 55 60

Ile Ile Gly Tyr Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Arg Gly Asn Pro Thr Val Glu Val Glu Val Tyr Thr Glu Thr Gly
1 5 10 15

Ala Phe Gly Arg Ala Leu Val Pro Ser Gly Ala Ser Thr Gly Gln Tyr
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Asp Arg Tyr Leu Gly Lys
35 40 45

Gly Val Leu Thr Ala Val Asn Asn Val Asn Glu Ile Ile Ala Pro Glu
50 55 60

Leu Leu Gly Phe Asp Val
65 70

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1511 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCCCGCGGCA ACCCAACACT TGAAGTAGAA GTTTACACTG AATCAGGTGC TTTCGGACGT	60
GGTATGGTTC CATCAGGAGC TTCTACTGGT GAACACGAAG CAGTTGAACT TCGCGACGGT	120
GACAAATCTC GTTACGGTGG TCTTGGTACA CAAAAGCTG TTGACAACGT AAACAACATC	180
ATTGCTGATT CTATCATTTG CTACGATGTA CGTGATCAAC AAGCTATTGA CCGTGCTATG	240
ATCGCACTTG ACGGTACTCC TAACAAAGGT AAATTGGGTG CGAATGCAAT CCTCGGTGTG	300
TCTATCGCTG TAGCTCGTGC TGCTGCTGAC TACCTTGAAA TCCCACTTTA CAGCTACCTT	360
GGTGGATTCA AACTAAAGT TCTTCCAAC CCAATGATGA ACATCATCAA CGGTGGTTCT	420
CACTCTGACG CTCCAATCGC TTTCCAAGAG TTCATGATCT TGCCAGTTGG TGCGCCAACA	480
TTTAAAGAAA CCCTTCGTTA CGGTGCTGAA ATCTTCCACG CTCTTAAGAA AATCCTTAAA	540
TCACGTGGTT TGGAAACTGC CGTAGGTGAC GAAGGTGGAT TCGCTCCTCG TTTCGAAGGA	600
ACTGAAGATG GTGTTGAAAC TATCCTTGCT GCGATTGAAG CTGCTGGATA TGTACCAGGT	660

(2) INFORMATION FOR SEQ ID NO:19:

(A) LENGTH: 420 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: B. subtilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Arg Gly Asn Pro Thr Leu Glu Val Glu Val Tyr Thr Glu Ser Gly
1 5 10 15

Ala Phe Gly Arg Gly Met Val Pro Ser Gly Ala Ser Thr Gly Glu His
20 25 30

Glu Ala Val Glu Leu Arg Asp Gly Asp Lys Ser Arg Tyr Gly Gly Leu
35 40 45

Gly Thr Gln Lys Ala Val Asp Asn Val Asn Asn Ile Ile Ala Asp Ser
50 55 60

Ile Ile Gly Tyr Asp Val Arg Asp Gln Gln Ala Ile Asp Arg Ala Met
65 70 75 80

Ile Ala Leu Asp Gly Thr Pro Asn Lys Gly Lys Leu Gly Ala Asn Ala
85 90 95

Ile Leu Gly Val Ser Ile Ala Val Ala Arg Ala Ala Ala Asp Tyr Leu
100 105 110

Glu Ile Pro Leu Tyr Ser Tyr Leu Gly Gly Phe Asn Thr Lys Val Leu
115 120 125

Gly Ala Ala Asn Ser Ile Leu Ile Lys Val Asn Gln Ile Gly Thr Leu

335

Asp Leu Lys Lys
420

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 429 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: S. pneumoniae

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAAAAAGGTA GCAGAAGCTG AGAAGAAGGT TGAAGAAGCT GAGAAAAAAG CCAAGGATCA 60
AAAAGAAGAA GATCGCCGTA ACTACCCAAC CAATACTTAC AAAACGCTTG ACCTTGAAAT 120
TGCTGAGTCC GATGTGAAAG TTAAAGAAGC GGAGCTTGAA CTAGTAAAAG AGGAAGCTAA 180
GGAACCTCGA GACGAGGAAA AAATTAAGCA AGCAAAAGCG AAAGTTGAGA GTAAAAAAGC 240
TGAGGCTACA AGGTTAGAAA ACATCAAGAC AGATCGTAAA AAAGCAGAAG AAGAAGCTAA 300
ACGAAAAGCA GCAGAAGAAG ATAAAGTTAA AGAAAAACCA GCTGAACAAC CACAACCAGC 360
GCCGGTTACT CAACCAGAAA AACCAGCTCC AAAACCAGAG AAGCCAGCTG AACAAACCAA 420
AGCAGAAAA 429

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *S. pneumoniae*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Lys Lys Val Ala Glu Ala Glu Lys Lys Val Glu Glu Ala Glu Lys Lys
 1 5 10 15

Ala Lys Asp Gln Lys Glu Glu Asp Arg Arg Asn Tyr Pro Thr Asn Thr
 20 25 30

Tyr Lys Thr Leu Asp Leu Glu Ile Ala Glu Ser Asp Val Lys Val Lys
 35 40 45

Glu Ala Glu Leu Glu Leu Val Lys Glu Glu Ala Lys Glu Pro Arg Asp
 50 55 60

Glu Glu Lys Ile Lys Gln Ala Lys Ala Lys Val Glu Ser Lys Lys Ala
 65 70 75 80

Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys Lys Ala Glu
 85 90 95

Glu Glu Ala Lys Arg Lys Ala Ala Glu Glu Asp Lys Val Lys Glu Lys
 100 105 110

Pro Ala Glu Gln Pro Gln Pro Ala Pro Val Thr Gln Pro Glu Lys Pro
 115 120 125

Ala Pro Lys Pro Glu Lys Pro Ala Glu Gln Pro Lys Ala Glu
 130 135 140

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTTNCTC GATGTCTCNG TNGCCAT

27

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCATAGTCT TCTTCGACTT GTTGATCATC

30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2462 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAAGCAGCC ACTTCTTCTA ATATGGCAAA GACAGAACAT AGGAAAGCTG CTAAACAAGC	60
GTTCGATGAAT ATATAGAAAA AATGATGAGG GAGATTCAAC TAGATAGAAG AAAACATACC	120
CAAAATGTCG CCTTAAACAT AAAGTTGAGC GCAATTAAAA CGAAGTATTT GCGTGAATTA	180
AATGTTTTAG AAGAGAAGTC GAAAGATGAG TTGCCGTCAG AAATAAAAGC AAAGTTAGAC	240
GCAGCTTTTG AGAAGTTTAA AAAAGATACA TTGAAACCAG GAGAAAAGGT AGCAGAAGCT	300
AAGAAGAAGG TTGAAGAAGC TAAGAAAAAA GCCGAGGATC AAAAAGAAGA AGATCGTCGT	360
AACTACCCAA CCAATACTTA CAAAACGCTT GAACTTGAAA TTGCTGAGTT CGATGTGAAA	420
GTTAAAGAAG CGGAGCTTGA ACTAGTAAAA GAGGAAGCTA AAGAATCTCG AAACGAGGGC	480
ACAATTAAGC AAGCAAAAGA GAAAGTTGAG AGTAAAAAAG CTGAGGCTAC AAGGTTAGAA	540
AACATCAAGA CAGATCGTAA AAAAGCAGAA GAAGAAGCTA AACGAAAAGC AGATGCTAAG	600
TTGAAGGAAG CTAATGTAGC GACTTCAGAT CAAGGTAAAC CAAAGGGGCG GGCAAAACGA	660

GGAGTTCCTG	GAGAGCTAGC	AACACCTGAT	AAAAAAGAAA	ATGATGCGAA	GTCTTCAGAT	720
TCTAGCGTAG	GTGAAGAAAC	TCTTCCAAGC	TCATCCCTGA	AATCAGGAAA	AAAGGTAGCA	780
GAAGCTGAGA	AGAAGGTTGA	AGAAGCTGAG	AAAAAAGCCA	AGGATCAAAA	AGAAGAAGAT	840
CGCCGTAAC	ACCCAACCAA	TACTTACAAA	ACGCTTGACC	TTGAAATTGC	TGAGTCCGAT	900
GTGAAAGTTA	AAGAAGCGGA	GCTTGAAC	GTAAAAGAGG	AAGCTAAGGA	ACCTCGAGAC	960
GAGGAAAAAA	TTAAGCAAGC	AAAAGCGAAA	GTTGAGAGTA	AAAAAGCTGA	GGCTACAAGG	1020
TTAGAAAACA	TCAAGACAGA	TCGTAAAAAA	GCAGAAGAAG	AAGCTAAACG	AAAAGCAGCA	1080
GAAGAAGATA	AAGTTAAAGA	AAAACCAGCT	GAACAACCAC	AACCAGCGCC	GGCTACTCAA	1140
CCAGAAAAAC	CAGCTCCAAA	ACCAGAGAAG	CCAGCTGAAC	AACCAAAAGC	AGAAAAAACA	1200
GATGATCAAC	AAGCTGAAGA	AGACTATGCT	CGTAGATCAG	AAGAAGAATA	TAATCGCTTG	1260
ACTCAACAGC	AACCGCCAAA	AACTGAAAAA	CCAGCACAAC	CATCTACTCC	AAAAACAGGC	1320
TGGAAACAAG	AAAACGGTAT	GTGGTACTTC	TACAATACTG	ATGGTTCAAT	GGCAACAGGA	1380
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1440
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAT	CTAAACGCTA	ATGGTTCAAT	GGCAACAGGA	1500
TGGCTCCAAA	ACAATGGTTC	ATGGTACTAC	CTAAACGCTA	ATGGTGCTAT	GGCGACAGGA	1560
TGGCTCCAAT	ACAATGGTTC	ATGGTACTAC	CTAAACAGCA	ATGGCGCTAT	GGCGACAGGA	1620
TGGCTCCAAT	ACAATGGCTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1680
TGGCTCCAAA	ACAACGGTTC	ATGGTACTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGA	1740
TGGCTCCAAT	ACAACGGTTC	ATGGTATTAC	CTCAACGCTA	ATGGTGATAT	GGCGACAGGT	1800

TGGGTGAAAG ATGGAGATAC CTGGTACTAT CTTGAAGCAT CAGGTGCTAT GAAAGCAAGC	1860
CAATGGTTCA AAGTATCAGA TAAATGGTAC TATGTCAATG GCTCAGGTGC CCTTGCAGTC	1920
AACACAACCTG TAGATGGCTA TGGAGTCAAT GCCAATGGTG AATGGGTAAA CTAAACCTAA	1980
TATAACTAGT TAATACTGAC TTCCTGTAAG AACTTTTTTAA AGTATTCCCT ACAAATACCA	2040
TATCCTTTCA GTAGATAATA TACCCTTGTA GGAAGTTTAG ATTAAAAAAT AACTCTGTAA	2100
TCTCTAGCCG GATTTATAGC GCTAGAGACT ACGGAGTTTT TTTGATGAGG AAAGAATGGC	2160
GGCATTCAAG AGACTCTTTA AGAGAGTTAC GGGTTTTTAA CTATTAAGCC TTCTCCAATT	2220
GCAAGAGGCT TCAATCTCTG CTAGGGTGCT AGCTTGCGAA ATGGCTCCAC GGAGTTTGGC	2280
AGCGCCAGAT GTTCCACGGA GATAGTGAGG AGCGAGGCCG CGGAATTCAC GAACTGCGAC	2340
GTTTTCTCCT TTGAGGTAA TCAATCGTTT CAAGTGTTCTG TAGGCGATCT TCATCTTGTC	2400
TTCAAAGGTC AAATCAGGTA GGATTCTCTC TGTTCAAAG TTTATGGTGG CCCTGGTTGA	2460
AG	2462

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

Met Met Arg Glu Ile Gln Leu Asp Arg Arg Lys His Thr Gln Asn Val
1 5 10 15

Leu Asn Val Leu Glu Glu Lys Ser Lys Asp Glu Leu Pro Ser Glu Ile
35 40 45

Lys Pro Gly Glu Lys Val Ala Glu Ala Lys Lys Lys Val Glu Glu Ala
65 70 75 80

Thr Asn Thr Tyr Lys Thr Leu Glu Leu Glu Ile Ala Glu Phe Asp Val
100 105 110

Ser Arg Asn Glu Gly Thr Ile Lys Gln Ala Lys Glu Lys Val Glu Ser
130 135 140

Lys Lys Ala Glu Ala Thr Arg Leu Glu Asn Ile Lys Thr Asp Arg Lys
145 150 155 160

Lys Ala Glu Glu Glu Ala Lys Arg Lys Ala Asp Ala Lys Leu Lys Glu
165 170 175

Lys Ala Glu Lys Thr Asp Asp Gln Gln Ala Glu Glu Asp Tyr Ala Arg

380

Trp Tyr Tyr Leu Asn Ala Asn Gly Asp Met Ala Thr Gly Trp Val Lys
565 570 575

Asn Gly Glu Trp Val Asn Glx
625 630